The Coupling of NSSL Warn-on-Forecast and FLASH Systems for Probabilistic Flash Flood Prediction

NUSRAT YUSSOUF

Cooperative Institute for Mesoscale Meteorological Studies, and School of Meteorology, University of Oklahoma, and NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma

KATIE A. WILSON, STEVEN M. MARTINAITIS, AND HUMBERTO VERGARA

Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, and NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma

PAMELA L. HEINSELMAN AND JONATHAN J. GOURLEY

NOAA/OAR/National Severe Storms Laboratory, and School of Meteorology, University of Oklahoma, Norman, Oklahoma

ABSTRACT

The goal of the National Oceanic and Atmospheric Administration's (NOAA) Warn-on-Forecast (WoF) program is to provide frequently updating, probabilistic model guidance that will enable National Weather Service (NWS) forecasters to produce more continuous communication of hazardous weather threats (e.g., heavy rainfall, flash floods, damaging wind, large hail, and tornadoes) between the watch and warning temporal and spatial scales. To evaluate the application of this WoF concept for probabilistic short-term flash flood prediction, the 0–3-h rainfall forecasts from NOAA National Severe Storms Laboratory's (NSSL) experimental WoF System (WoFS) were integrated as the forcing to the NWS operational hydrologic modeling core within the Flooded Locations and Simulated Hydrographs (FLASH) system. Initial assessment of the potential impacts of probabilistic short-term flash flood forecasts from this coupled atmosphere–hydrology (WoFS-FLASH) modeling system were evaluated in the 2018 Hydrometeorology Testbed Multi-Radar Multi-Sensor Hydrology experiment held in Norman, Oklahoma. During the 3-week experiment period, a total of nine NWS forecasters analyzed three retrospective flash flood events in archive mode. This study will describe specifically what information participants extracted from the WoFS-FLASH products during these three archived events, and how this type of information is expected to impact operational decision-making processes. Overall feedback from the testbed participants' evaluations show promise for the coupled NSSL WoFS-FLASH system probabilistic flash flood model guidance to enable earlier assessment and detection of flash flood threats and to advance the current warning lead time for these events.

1. Introduction

The National Oceanic and Atmospheric Administration (NOAA) National Severe Storms Laboratory's (NSSL) Warn-on-Forecast program (WoF; Stensrud et al. 2009, 2013) is developing a frequently cycled, probabilistic, convective-scale, numerical weather prediction (NWP) model-based ensemble system, which is referred to as the Warn-on-Forecast System (WoFS). The vision of the WoF program is to fill the gap in forecasters' current watch-to-warning paradigm for severe thunderstorm, tornadoes, heavy rainfall, flash floods and other

hazardous weather, where guidance from NWP models currently play a less significant role. The watch typically covers an area of about $65\,000\,\mathrm{km}^2$ and issued for a duration of approximately 4–8 h whereas the warning typically covers about $1500\,\mathrm{km}^2$ for a duration of ~1–2 h. The frequently updated WoFS ensembles are postprocessed to provide probabilistic forecast guidance, which is anticipated to enhance forecasters' abilities to provide a more continuous flow of probabilistic forecasts for high-impact weather between the watch and warning spatial and temporal scales. This guidance supports the concepts of the Forecasting a Continuum of Environmental Threats (FACETs) program (Rothfusz et al. 2018), which aims to modernize the current

Corresponding author: Nusrat Yussouf, nusrat.yussouf@noaa.gov

National Weather Service (NWS) watch and warning system with a more continuous flow of probabilistic hazard forecasts on increasingly fine spatial and temporal scales. This modernization is expected to better support weather-related decisions for a variety of end users.

Several recent studies demonstrate the potential of the WoFS in forecasting skillful 0–3-h heavy rainfall with reasonable accuracy in areal coverage and amount (Yussouf et al. 2016; Lawson et al. 2018; Yussouf and Knopfmeier 2019). While severe thunderstorm and tornado forecasts depend on modeling of atmospheric variables (Wheatley et al. 2015; Yussouf et al. 2015; Skinner et al. 2016, 2018), flash flooding forecasts depend on both atmospheric and hydrological conditions (Doswell et al. 1996; Davis 2001; Sorooshian et al. 2008). The hydrologic response of the watershed where heavy rainfall accumulates must be considered when forecasting flash flooding. It is therefore critical to integrate the quantitative precipitation forecasts (QPFs) from an atmospheric model as a forcing to a distributed hydrologic model for generating surface water and routing stream discharge products to produce explicit flash flood forecasts (Brown et al. 2012; Vincendon et al. 2011; Hardy et al. 2016; Amengual et al. 2017). The current state-of-the-art distributed hydrologic models use rainfall observations as the forcing mechanism to predict deterministic stream discharge products (Devi et al. 2015). One example of this is the NSSL Flooded Locations and Simulated Hydrographs (FLASH; Gourley et al. 2017) system that generates deterministic stream discharge products using the Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimate (QPE) as a forcing mechanism (Zhang et al. 2016). The successful evaluation of the deterministic flash flood products from the FLASH system during the 2014–16 MRMS Hydrometeorology Testbed–MRMS Hydrology (HMT-Hydro) experiment (Martinaitis et al. 2017) resulted in the transition of the FLASH suite to the National Centers for Environmental Prediction (NCEP) in 2018. The FLASH products are used routinely in NWS weather forecast offices and at the NWS Weather Prediction Center for flash flood monitoring, detection, and decision making. However, the operational FLASH products neither provide probabilistic information to communicate the uncertainty associated with the forecast, nor do they provide significantly longer flash flood forecast lead time. Therefore, to extend the hydrometeorological forecast lead time beyond the watershed response time, it is prudent to explore the use of short-term quantitative precipitation forecasts (QPFs) from NWP models as a forcing to the hydrologic model (Hardy et al. 2016, and references therein). Given the

ability of the WoFS to predict high-resolution, short-term, convectively driven rainfall forecasts with similar spatial and temporal scale modeling in the FLASH system, the QPFs and probabilistic QPFs (PQPFs) from the WoFS can be used as a forcing pathway for extending lead times in flash flood forecasting.

The potential application of WoFS 0–3-h ensemble QPFs for flash flood prediction was analyzed during the 2018 HMT-Hydro experiment that was held in Norman, Oklahoma (OK). This experiment was the first attempt to couple an atmospheric and a hydrologic ensemble system for probabilistic flash flood forecasts at the storm scale. This system is referred to as the WoFS QPF-forced FLASH system (WoFS-FLASH). The HMT-Hydro experiment ran for 3 weeks, starting on 25 June and ending on 20 July 2018 with a 1-week break in between. A total of nine participants from the NWS participated in the evaluation activities. Activities each week consisted of 2 days of real-time experimental warning operations and 2 days of WoFS case study analysis. Three retrospective flash flood events were analyzed during the 2 days of case study activities to assess the potential operational impacts of ingesting WoFS QPFs into the flash flood prediction process. The testbed participants assessed flash flood threats for both deterministic and probabilistic flash flood products using three data conditions: 1) deterministic flash flood products from QPE-forced FLASH system, 2) probabilistic flash flood products from QPE-forced FLASH, and 3) probabilistic flash flood guidance from WoFS-FLASH. Comparisons between participants' assessments of the flash flood threat for the three data conditions were qualitatively analyzed to assess how each data condition contributed to participants' interpretation and expectations for flash flooding events. Additionally, on the last day of each experiment week, participants contributed to a focus group discussion to share overall impressions on the usefulness of the probabilistic forecast guidance for flash flood decisions making. This paper will describe a first assessment for how the WoFS-FLASH modeling system performed, and how forecasters envision the resulting products will impact their understanding and subsequent decision making for flash flooding events.

2. Methods

a. The experimental WoFS

The experimental WoFS (Yussouf et al. 2015, 2016; Wheatley et al. 2015; Jones et al. 2016; Yussouf and Knopfmeier 2019) used in this study is a 36-member Advanced Research version of the Weather Research and Forecasting (WRF-ARW; Skamarock et al. 2008)

FIG. 1. Schematic of the experiment design: (a) MRMS QPE and WoF QPF from the three archived cases integrated into (b) the FLASH EF5 system and (c) the probabilistic flash flood products are evaluated at the 2018 HMT-Hydro experiment.

Model-based ensemble data assimilation and prediction system. The WoFS domain covered a ~900-km wide region at 3-km horizontal grid spacing and was centered over the region where the hazardous weather was anticipated. There are 51 vertical grid levels that extend from the surface to 10 hPa at the top. All ensemble members utilized the NSSL 2-moment microphysics parameterization and the Rapid Refresh (RAP) land surface model, but the planetary boundary layer (PBL) and radiation physics options were varied among the ensemble members to address uncertainties in model physics. The WoFS used the experimental High-Resolution Rapid Refresh (HRRR) ensemble (HRRRE; Dowell et al. 2016) developed by Earth System Research Laboratory's Global System Division for initial and boundary conditions. Once initialized from the HRRRE, the WoFS was cycled every 15 min to assimilate all available storm observations into the system, including MRMS radar reflectivity (Smith et al. 2016) and Level II radial velocity data, cloud water path retrievals from the satellite, and conventional observations (e.g., Meteorological Aerodrome Reports, Automated Surface Observing Systems, radiosonde, aircraft, marine, and mesonet). Observations were assimilated using the Community Gridpoint Statistical Interpolation (GSI; DTC 2017a), ensemble Kalman filter (EnKF; Whitaker et al. 2008; DTC 2017b) data assimilation (DA) technique and thus provided the capability to update the ensemble probabilistic model guidance on a 15-min basis. For each of the three flash flood events, the

0–3-h WoFS forecasts were generated at the top of each hour for the duration of the evaluation period. The WoFS ensemble files were created at 5-min intervals and the rainfall forecasts from each of the ensemble members were used as a forcing to drive the experimental probabilistic FLASH system (Figs. 1a,b).

b. The FLASH system

FLASH's hydrologic modeling core—referred to as the Ensemble Framework for Flash Flood Forecasting (EF5)—was employed for the forward simulation of surface flow rates. EF5 features multiple physical representations of the rainfall–runoff process to generate products relevant to the occurrence of flash floods including maxima of streamflow, unit streamflow (streamflow normalized by upstream drainage area), and soil saturation. For this particular study, an adaptation of the water balance component of the Coupled Routing and Excess Storage (CREST; Wang et al. 2011) and the kinematic wave approximation of the Saint-Venant equations of one-dimensional open channel flow within EF5 were used.

MRMS radar-only QPE is used to force all three data conditions up to analysis time to update the hydrologic model's initial conditions for each new forecast. For conditions 1 and 2, no precipitation is available to the hydrologic model beyond analysis time such that the flash flood forecast is only based on the hydrologic response to past precipitation. However, for condition 3, the 3-h rainfall forecast from WoFS is used to force the

FLASH system out to 3 h beyond the analysis time for conditions 1 and 2. The flash flood forecast for condition 3 is based on the hydrologic response beyond the third hour.

The flash flood forecast cycling for all experiments and conditions mimicked that of the operational FLASH system: a new hydrologic forecast out to 12 h at a 5-min time step was launched every 10 min and produced maxima of streamflow, unit streamflow, and soil saturation. The resulting deterministic products (condition 1) are input to a postprocessing algorithm based on the statistical analysis of conditional distributions of measured streamflow to generate the probabilistic FLASH forecasts in condition 2. Condition 3 utilized the 36 individual WoFS ensemble members to generate probabilistic FLASH forecasts. All 36 ensemble member WoFS QPFs are ingested in the EF5, which results in 36 forecasts of maximum unit streamflow [described in section 2c(4)]. Each of these 36 maximum unit streamflow forecasts goes into the postprocessing algorithm to produce a singular mean output of the different probabilistic FLASH products. The probabilistic product suite from FLASH using WoFS was then placed into the Advanced Weather Interactive Processing System (AWIPS) in a similar format to accessing NWP model parameters and forecasts within the user interface.

c. Deterministic and probabilistic gridded flash flood products

The testbed participants evaluated a series of QPE comparison products as well as gridded deterministic and probabilistic flash flood products from the FLASH hydrologic system for each of the three archived case studies. The gridded flash flood products were created from the CREST model at $0.01° \times 0.01°$ (approximately $1 \text{ km} \times 1 \text{ km}$) grid spacing and at a 10-min temporal resolution. The baseline products consisted of MRMS reflectivity, radar-only QPE, and QPE comparisons such as the QPE-to-FFG (flash flood guidance) ratio and QPE average recurrence interval. The maximum unit streamflow from the QPE-FLASH system is used as the benchmark deterministic hydrologic model product. To evaluate the probability of flash flood occurrence, four experimental exceedance probability products were used. These include the probability of receiving a flash flood report and the probability of exceeding maximum unit streamflow values to determine minor, moderate, and major flash flooding potentials. These probabilistic products were used to compare the QPE-forced FLASH to the WoFS-FLASH system. A depiction of the primary products used to evaluate the relative

performance of the three different conditions is provided in Figs. 2a–h.

1) MRMS RADAR-ONLY QPE

The MRMS system uses radar, surface, and satellite observations, NWP models, and precipitation climatologies to generate very high resolution QPE (Zhang et al. 2016). This study utilized the radar-only QPE instantaneous rain rates. These rates are generated by the mosaicking of quality-controlled radar reflectivity that mitigate the impacts of nonmeteorological echoes, bright banding, beam blockages, and other artifacts. These radars are seamlessly blended based on the weighting of various vertical reflectivity data with respect to the ground level and the freezing level. The final seamless reflectivity mosaic is then translated into precipitation rates based on a surface precipitation type classification scheme allowing for a unique reflectivity–rain rate relationship to be applied to each grid cell. The 2-min radar-based QPE products are aggregated to 1-h accumulation period (Fig. 2a). Depending on the users' need, the MRMS QPE products can be aggregated to longer accumulation time periods (e.g., 3, 6, 12, and 24 h) and these products are used in operations by the NWS and other government agencies.

2) QPE-TO-FFG RATIO

The QPE-to-FFG ratio shows the magnitude to which the estimated rainfall may exceed the amount of rainfall required to bring the river up to, and possibly beyond, bankfull conditions (Fig. 2b). This product compares the MRMS radar-only QPEs to corresponding FFG (Sweeney 1992; Clark et al. 2014) values at the MRMS spatiotemporal resolution (Gourley et al. 2017). FFG values are generated at each NWS river forecast center (RFC) which is then mosaicked and delivered by Weather Prediction Center (WPC). There are no local NWS office enhancements of the FFG values accounted for in this FFG field. This ratio is calculated for 1-, 3-, and 6-h rainfall accumulation periods.

3) QPE AVERAGE RECURRENCE INTERVAL

The QPE average recurrence interval (ARI) product is used to identify the potential rarity of rainfall in a given location. It is a comparison of the MRMS radar-only QPE to the NOAA Atlas 14 precipitation frequency values (Perica et al. 2013). Greater return periods indicate more significant events. Although the product does not have a direct relationship with flash flooding, it can provide an estimate on the rarity of the precipitation event. This product is calculated for 30-min and 1-, 3-, 6-, 12-, and 24-h accumulation periods
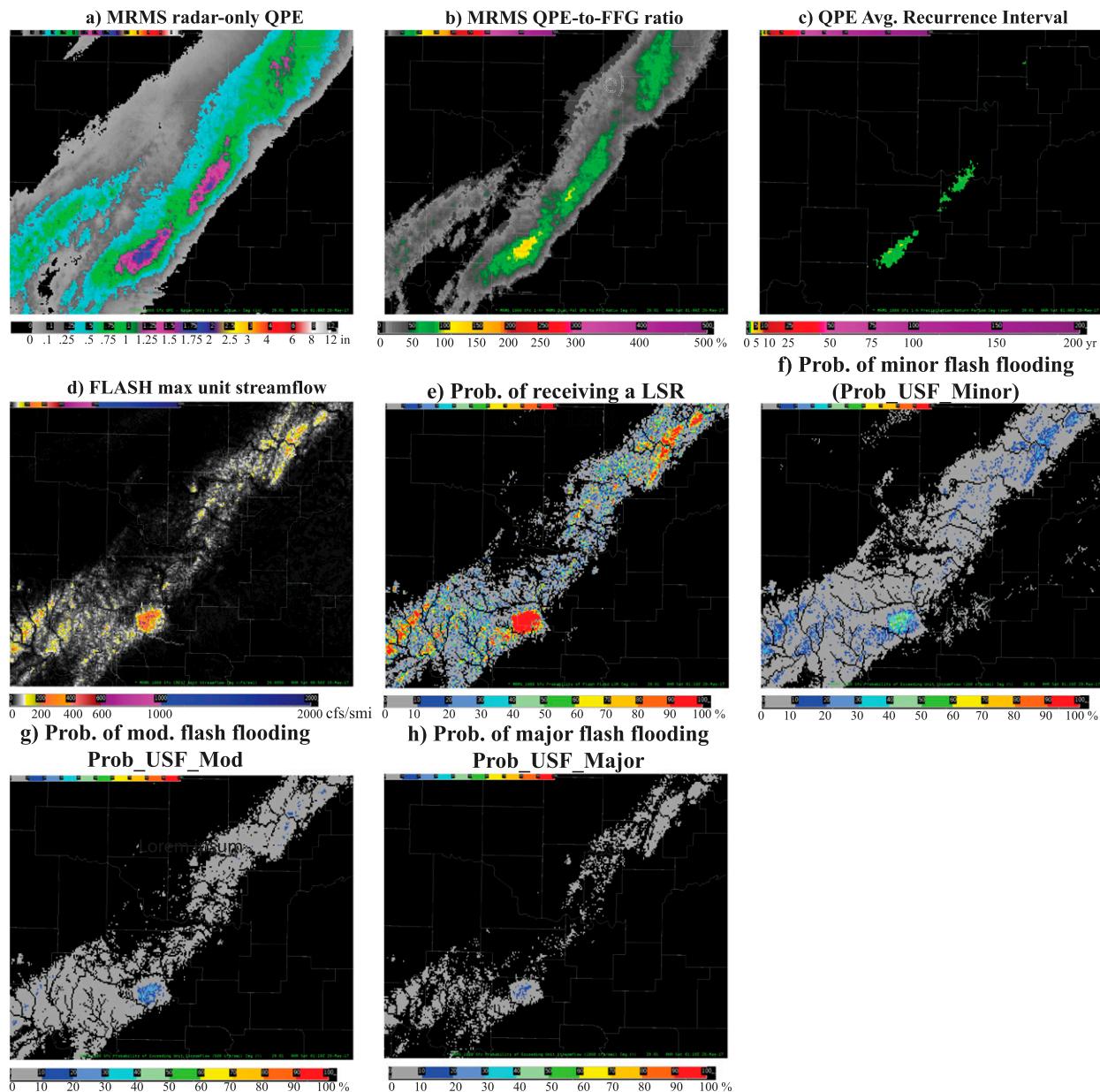
FIG. 2. The (a) 1-h accumulated MRMS radar-only QPE (in.), (b) 1-h MRMS QPE-to-FFG ratio (%), (c) QPE average recurrence interval (yr), (d) FLASH maximum unit streamflow ($m^3\,s^{-1}\,km^{-2}$), (e) probability of receiving a LSR (Prob_LSR), and probability of exceeding (f) $2\,m^3\,s^{-1}\,km^{-2}$ unit streamflow (Prob_USF_Minor; minor flash flood), (g) $5\,m^3\,s^{-1}\,km^{-2}$ unit streamflow (Prob_USF_Mod; moderate flash flood), and (h) $10\,m^3\,s^{-1}\,km^{-2}$ unit streamflow (Prob_USF_Major; major flash flood). The plots shown here are from the Murray and Carter Counties, OK, flash flood event (valid 0100 UTC 20 May 2017).

and is also generated at the MRMS spatiotemporal resolution (Fig. 2c).

4) MAXIMUM UNIT STREAMFLOW

Maximum unit streamflow is a deterministic hydrologic product that is used to diagnose areas of flash flooding potential, as well as to identify the relative severity of the potential flash flooding impacts (Fig. 2d;

Gourley et al. 2017; Martinaitis et al. 2017). The maximum unit streamflow ($m^3\,s^{-1}\,km^{-2}$) is computed by taking the maximum streamflow throughout the forecast period and then dividing this number by the upstream drainage area. An area of contiguous grid points with high values of maximum unit streamflow is usually a cause for concern for flash flooding threat. This product can be used to visualize stream and river

networks and to identify broad areas where land surface areas are being inundated.

5) PROBABILITY OF RECEIVING A FLASH FLOOD LOCAL STORM REPORT

The probability of receiving a flash flood local storm report (LSR; Prob_LSR) is based on a statistical analysis of flash flood reports from the NWS *Storm Data* and their associated CREST maximum unit streamflow values (Fig. 2e). An area of contiguous grid points with high values of Prob_LSR can be used to diagnose areas of flash flooding potential. A single or a handful of isolated grid points with higher probability values may not be indicative of a flash flooding threat.

6) PROBABILITY OF EXCEEDING MAXIMUM UNIT STREAMFLOW VALUE

The exceedance probabilities for three maximum unit streamflow thresholds are calculated and used as proxies for flash flood severity ranging from minor to major flash flooding. These threshold values are selected based on statistical analysis and derived probability distribution functions. The probabilities of exceeding maximum unit streamflow of $2\,\mathrm{m^3\,s^{-1}\,km^{-2}}$ $(183\,\mathrm{ft^3\,s^{-1}\,mi^{-2}})$, $5\,\mathrm{m^3\,s^{-1}\,km^{-2}}$ $(457.5\,\mathrm{ft^3\,s^{-1}\,mi^{-2}})$, and $10\,\mathrm{m^3\,s^{-1}\,km^{-2}}$ $(915\,\mathrm{ft^3\,s^{-1}\,mi^{-2}})$ are used as proxies for minor flash flooding (Prob_USF_Minor; Fig. 2e), moderate flash flooding (Prob_USF_Mod; Fig. 2f), and major flash flooding (Prob_USF_Major; Fig. 2g), respectively.

d. Forecaster evaluations

Nine NWS forecasters viewed the flash flood forecast products for three cases (described in section 3) in playback mode using the Weather Event Simulator (WES) in AWIPS-II. These products were viewed through a series of three data conditions. Condition 1 was based on deterministic flash flood forecast guidance from the QPE-FLASH system, condition 2 was based on probabilistic forecast guidance using QPE-forced FLASH, and condition 3 was based on the probabilistic forecast guidance using WoFS-FLASH. Products made available in condition 1 are currently used in NWS operations and therefore provided a reference benchmark to evaluate the impact of conditions 2 and 3. Guidance was valid only at the top of the hour for conditions 1 and 2, while forecast guidance was valid at the top of the hour and out to the next 3 h for condition 3. Working independently and at their own pace, participants cycled through viewing the three data conditions and assessed the potential flash flood threat.

A data collection tool was used to guide participants' evaluations of the three data conditions during every hour of the weather scenario. For both conditions 1 and 2, participants were asked three questions: 1) what is your current understanding of the flash flood threat, 2) what information did you extract from these products, and 3) would you have taken any action (by issuing an advisory, warning, follow-up statement, etc.) at this time based on these products? These same questions were asked in condition 3, however participants were asked to focus on times within the 3-h forecast that were of most interest to them. Participants were encouraged to use county names and/or cities to identify different threat areas throughout their evaluations. Evaluations provided qualitative descriptions only and did not include the actual issuance of operational products (e.g., warnings).

Evaluations were first analyzed with respect to 1) where participants chose to focus their attention (at a county-level resolution) and 2) what their expected actions were. Participants' records of what locations they were attending to and their expected actions captured overall similarities and differences between how they viewed and interpreted the flash flood products while advancing through the three data conditions. The information participants used to direct attention and expected actions in the three data conditions is discussed for each weather scenario.

3. Description of the flash flood events

a. Case 1: 19–20 May 2017 flash flood in Murray and Carter Counties, Oklahoma

The area of interest for the 19–20 May 2017 event focused on Murray and Carter Counties in south-central OK (Fig. 3a). At around 1900 Coordinated Universal Time (UTC), convection initiated in north Texas and moved toward the northeast into south-central OK. The WPC Mesoscale Precipitation Discussions (MPD) highlighted the potential for 2–4 in. of rain in that area. The first flash flood was reported in Murray County, OK, at 0155 UTC 20 May 2017 per NWS *Storm Data*. Numerous roads and state highways were flooded around the cities of Davis and Sulphur in Murray County and Pooleville in Carter County, OK, along with a report of water entering the basement of a local elementary school. Falls Creek in Murray County rose approximately 9 ft in 1 h in response to the heavy rainfall, which flooded all bridges at a local campground. Participants focused on the time period between 2000 UTC 19 May and 0300 UTC 20 May 2017 during the evaluation.
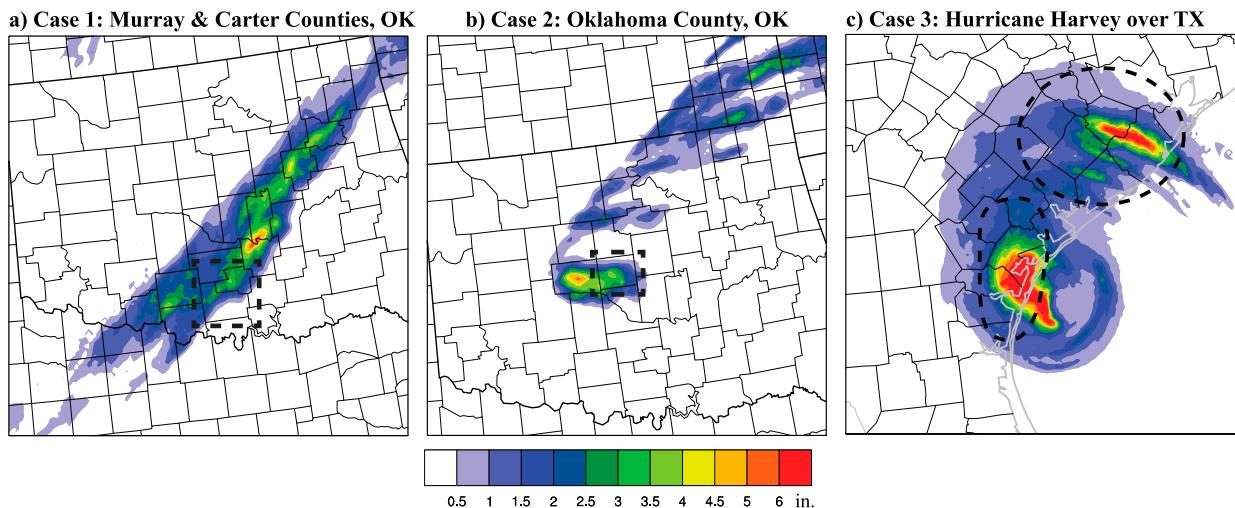
FIG. 3. NCEP's Stage IV 6-h rainfall totals (in.) valid at (a) 0200 UTC 20 May 2017, (b) 0200 UTC 1 Jun 2013, and (c) 0400 UTC 26 Aug 2017. The dashed black outline overlaid on each panel is the location of interest for the flash flood event.

b. Case 2: 31 May–1 June 2013 flash flood over Oklahoma County, Oklahoma

A cluster of storms formed along a cold front in west-central Oklahoma around 2130 UTC. The storm cell that spawned the EF3 (on the Enhanced Fujita scale) El Reno tornado in between 2303 and 2343 UTC moved slowly toward the east while new convective cells regenerated near the original initiation point. This back-building storm system brought heavy rainfall to the Oklahoma City (OKC) metropolitan area, resulting in significant flash flooding during the evening of 31 May and the early morning of 1 June 2013 (Fig. 3b). The first flash flood was reported at 0100 UTC 1 June 2013 in Oklahoma County. A total of 13 people were killed by flash floods, including 12 people in OKC, making this event the deadliest flash flood event in OKC history (NWS 2015). Participants focused on the time period from 2300 UTC 31 May to 0100 UTC 1 June 2013.

c. Case 3: 26 August 2017 flash flood from Hurricane Harvey over Texas

Hurricane Harvey made landfall in the middle of the Texas Gulf Coast as a category 4 hurricane at 0300 UTC 26 August 2017. Preceding and during landfall, the well-formed rainband northeast of the eyewall positioned itself over southeastern Texas and produced significant rainfall. Harvey slowed considerably after landfall, stalled over southeast Texas for several days, and produced over 50 in. of rainfall resulting in historic and deadly flash flooding. The torrential tropical rains impacted the coastal counties of the Coastal Bend, as well as the Victoria Crossroads region, and several flash flood warnings (FFWs) were issued during the evening and

overnight hours. The participants focused on the time period between 0000 and 0700 UTC on 26 August 2017 and evaluated the flash flood products for two main locations: the northern area where the outer rainbands came onshore, and the southern area where the eyewall made landfall (Fig. 3c).

Note that the areas of highest rainfall do not necessarily coincide with areas of greatest flash flood impact. Heavy rainfall is more likely to result in flash flooding when it occurs in a flash flood–prone area or basin.

4. Results and discussions

a. Overall WoFS forecast quality

The WoFS ensemble probabilities of 0–3-h rainfall forecasts exceeding 25.4 mm (1.0 in.) for the three cases were calculated at model grid points to highlight areas of the most intense rainfall (Fig. 4). The 3-h rainfall totals generated 100% probabilities for the dominant precipitation core at the observed NCEP Stage-IV rainfall area with reasonable accuracy for all three cases. Most of the NWS *Storm Data* flash flood reports valid within the 3-h forecast time period were within the ensemble probability envelope for case 1 and case 2 (Figs. 4 a,b). The location and timing of the flash flood reports give an idea of the areal coverage and timing where heavy rainfall led to flash flooding. Even though the flash flood reports from NWS *Storm Data* are useful for validating ensemble-derived forecast products, subjectivity due to the human element in the flash flood reporting process must be taken into consideration when interpreting the results. There were no flash flood reports during the forecast time period for case 3 (Fig. 4c). There are a few
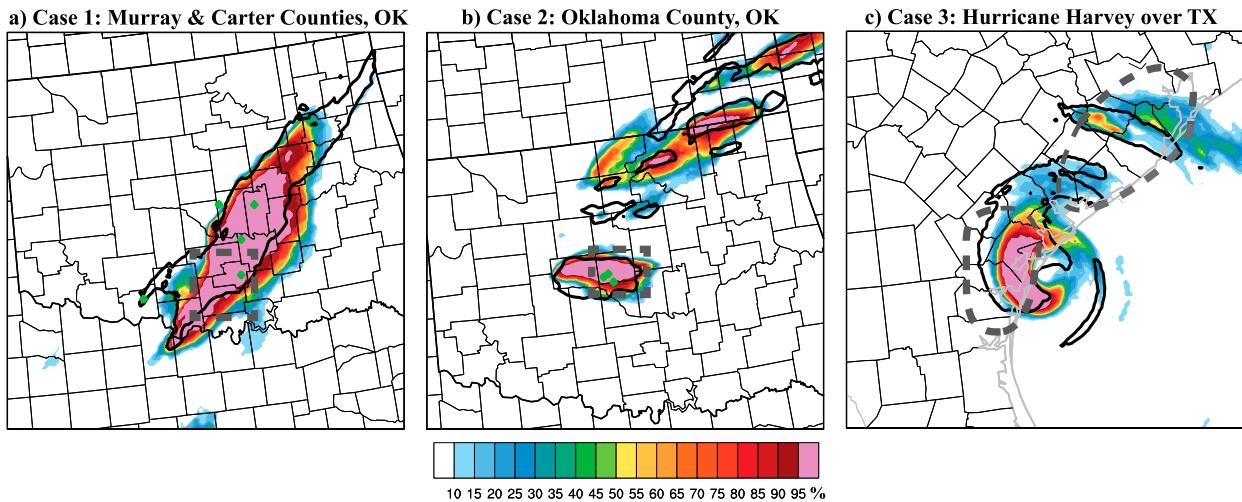
FIG. 4. Ensemble-derived probability of 0–3-h rainfall forecasts greater than 1.0 in. (colors; 5% increment) valid at (a) 0200 UTC 20 May 2017, (b) 0200 UTC 1 Jun 2013, and (c) 0400 UTC 26 Aug 2017. The thick black contours overlaid are the Stage-IV 1.0-in. rainfall, and the dashed gray outline overlaid on each panel is the location of interest for the flash flood event. The green dots are flash flood reports from NWS *Storm Data* during the forecast time period.

areas where the WoFS indicates no rainfall forecast in the observed location, for example, to the north of the high rainfall core for the OKC flash flood event (Fig. 4b). While rainfall probabilities were small over the outer rainband (Fig. 4c) for Hurricane Harvey, the system was able to capture the heavy rainfall generated from the eyewall convection. The ensemble fractions skill score (eFSS; Duc et al. 2013; Roberts and Lean 2008) is computed for 0–3-h WoFS rainfall using neighborhood widths of 0, 9, 18, and 27 km (Fig. 5) and Stage IV rainfall as the observations. The eFSS is a quantitative probabilistic verification measure of the spatial skill with a score of 1 indicating perfect skill. As the neighborhood increases, more ensemble members overlap or agree, which led to higher eFSS scores from 0 to 27 km. The eFSSs are higher than 0.5 for the three events at native grid points (0-km neighborhood) for all thresholds. Not

surprisingly, the eFSSs for all three events are high for small rainfall thresholds and the value decreases as the threshold increases.

b. Forecaster attention and expected actions

1) CASE 1: 19–20 MAY 2017

In the first hour of the case (2000 UTC), participants' monitoring of the weather scenario altered notably after viewing condition 3. Unlike in conditions 1 and 2, the availability of 3-h probabilistic flash flood forecasts resulted in seven participants attending to Murray/Carter County locations (Figs. 6 and 7). It was not until the third hour of the case (2200 UTC) when all data conditions led at least some participants to express concern for the flash flooding threat in Murray/Carter Counties. After viewing condition 1 data at 2200 UTC, two participants
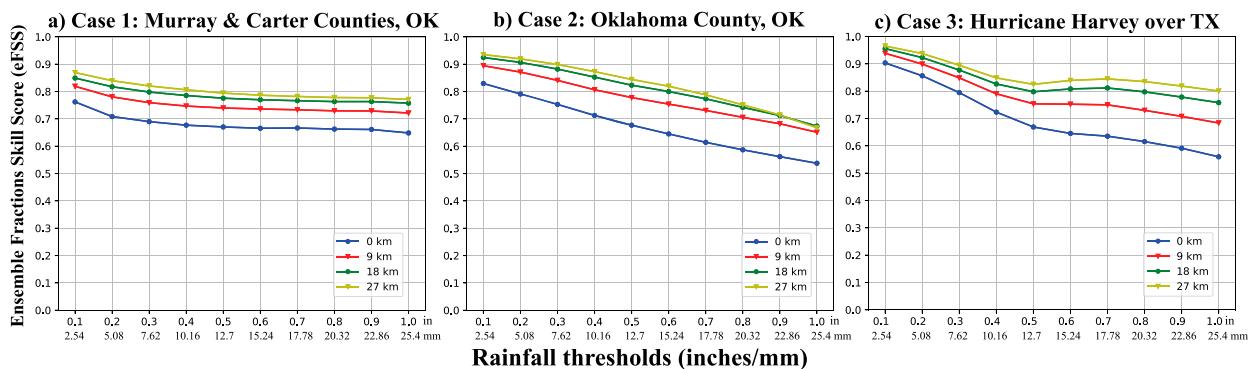

FIG. 5. Ensemble fractions skill score (eFSS) as a function of rainfall thresholds (in. or mm) for 0–3-h rainfall ensemble forecasts valid at (a) 0200 UTC 20 May 2017, (b) 0200 UTC 1 Jun 2013, and (c) 0400 UTC 26 Aug 2017. Details are shown in the legend.
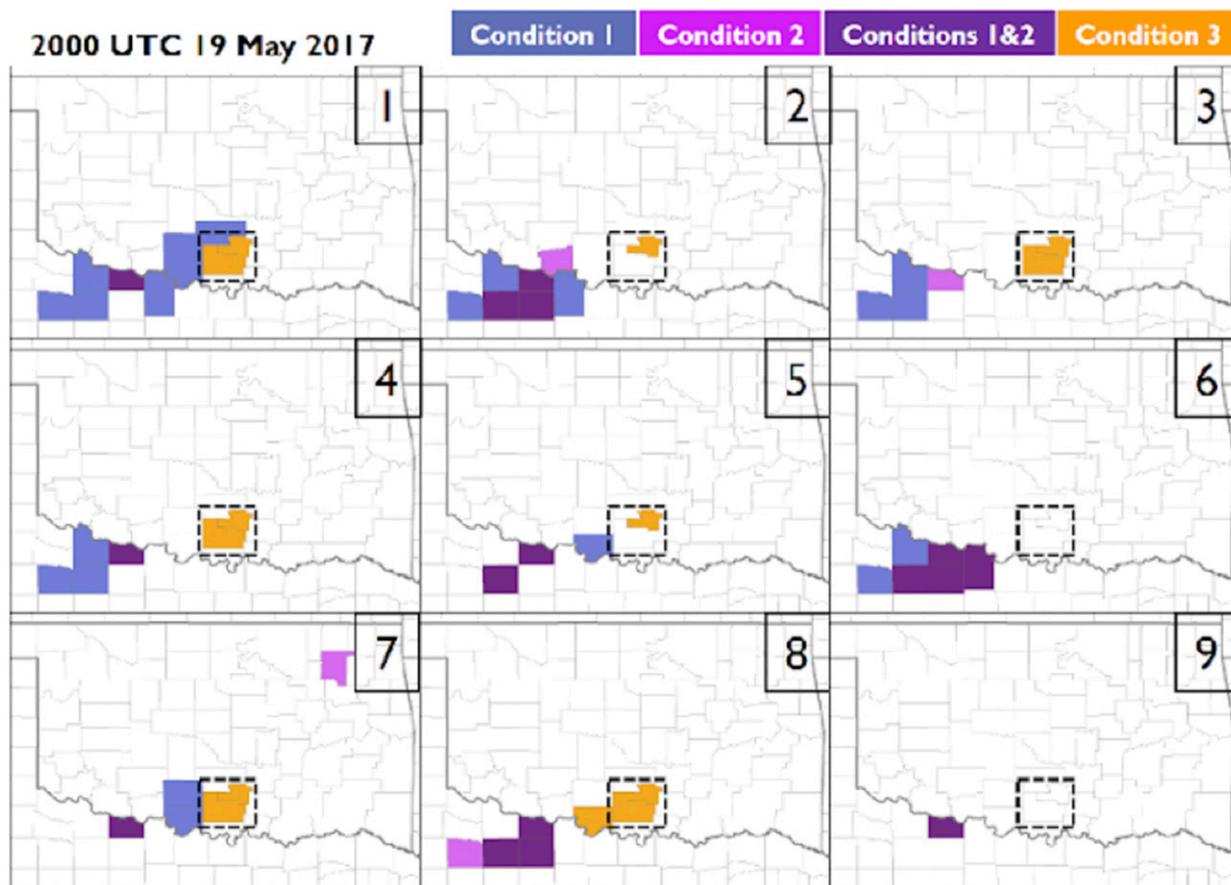
FIG. 6. Case 1: Anticipated participants' actions based on analysis of (top) condition 1, (middle) condition 2, and (bottom) condition 3 data at the top of each hour from 2000 UTC 19 May to 0300 UTC 20 May 2017 for the area highlighted in Fig. 5.

monitored the potential threat in this location, while another also communicated the threat due to the increasing signal in unit streamflow (Fig. 7). However, when provided with the probabilistic products in condition 2 for 2200 UTC, seven participants noticed very high Prob_LSR reaching 100% and low Prob_USF_Minor values of approximately 20%. This guidance resulted in three participants stating they would have issued a flash flood advisory and two participants stating they would have communicated the threat via NWSChat (Fig. 7). The 2200 UTC guidance in condition 3 led to a growing concern for flash flooding in Murray/Carter Counties compared to earlier forecasts after most participants observed increasing trends in Prob_USF_Minor values to 50% and Prob_USF_Mod values to 30%. Eight of the nine participants acknowledged this potential threat, resulting in expected actions that included two short-fused FFAs (flash flood watches), three FFWs (one of which escalated to a "flash flood emergency"), and three communications of the threat to emergency managers and/or the public (Fig. 7).

When viewing condition 1 over the next 2 h (2300–0000 UTC), only few participants noted a signal in the unit streamflow product in Murray/Carter Counties. While two participants reported they would issue a FFW due to heavy rain moving through the area, most participants' attention was elsewhere (Fig. 7). While viewing condition 2 for these same hours, flash flood advisories from previous decisions continued, but participants did not pay specific attention to Murray/Carter Counties. However, all participants paid attention to this location when viewing condition 3 data for the 2200 UTC and 2300 UTC forecasts. Participants reported impressive values in the Prob_LSR, Prob_USF_Minor, Prob_USF_Mod, and Prob_USF_Major products (Fig. 8), resulting in two additional participants reporting they would have issued a FFW (Fig. 7). Despite noticing increasing probability values in condition 3 compared to earlier in the case, three participants did not expect to take action beyond monitoring the threat. One of these participants questioned why the probability values declined during
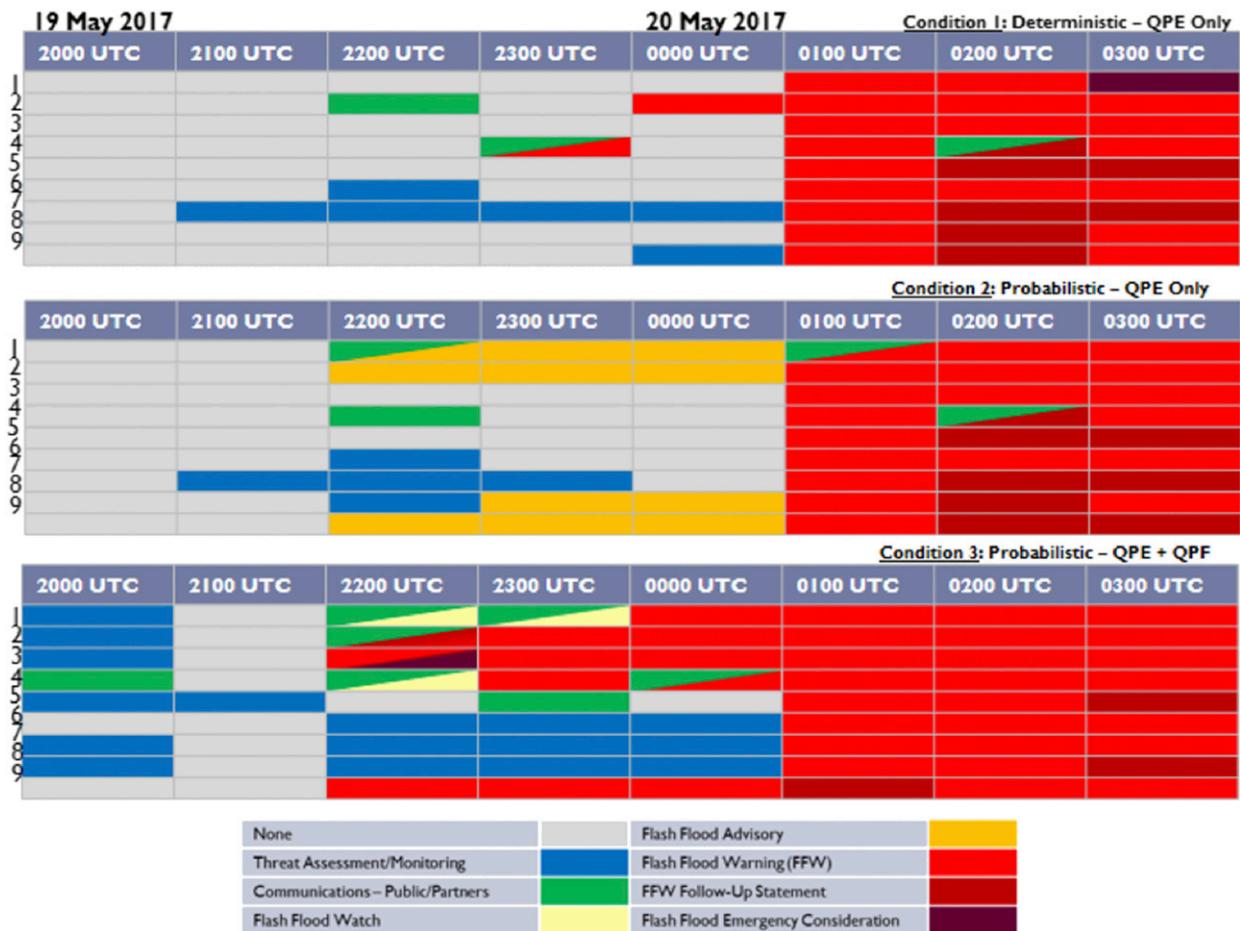
FIG. 7. Case 1: Anticipated participants' actions based on analysis of (top) condition 1, (middle) condition 2, and (bottom) condition 3 data at the top of each hour from 2000 UTC 19 May to 0300 UTC 20 May 2017 for the area highlighted in Fig. 6.

the third hour of WoFS-FLASH guidance (Fig. 8), while another questioned the validity of the high probability values.

By 0100 UTC, the flash flooding threat in Murray/Carter Counties was apparent in all three data conditions. It is also during this hour that the first flash flooding report was received (0155 UTC). Participants that had not already decided that a FFW was warranted did so when evaluating condition 1. Participants noted "quick developing heavy rains" and a "sudden jump" in unit streamflow values exceeding $6 \, \text{m}^3 \, \text{s}^{-1} \, \text{km}^{-2}$ ($549 \, \text{ft}^3 \, \text{s}^{-1} \, \text{mi}^{-2}$) in some areas. In condition 2 at 0100 UTC, participants maintained their expected action of FFW issuance but with lower confidence due to the probabilities being lower than what condition 1 led them to anticipate (e.g., Prob_USF_minor at 50%). After viewing condition 3 for this same hour, four participants noted that the flash flooding threat was worst for the Murray/Carter County locations during the 3-h forecast, and consequently expected they

would issue FFWs (Fig. 7). The remaining five participants instead focused their attention where the storm system was moving into ahead of Murray/Carter Counties.

In the final 2 h of the case (0200–0300 UTC), after seeing increasing values in maximum unit streamflow in condition 1 (Fig. 7), seven participants expected they would have either issued updates to FFWs, communicated the threat, or considered a flash flood emergency. Anticipating that the weather conditions would further deteriorate, six participants intended to take these actions primarily to elevate wording in messages. The probabilities viewed in condition 2 during these last 2 h affirmed the need for most participants' expected actions listed in condition 1. Unlike in conditions 1 and 2 however, participants' attention in condition 3 during 0200–0300 UTC was focused on the eastward expansion of the event. This shift in attention resulted in only four of nine participants mentioning the threat in Murray/Carter Counties during these hours, at which
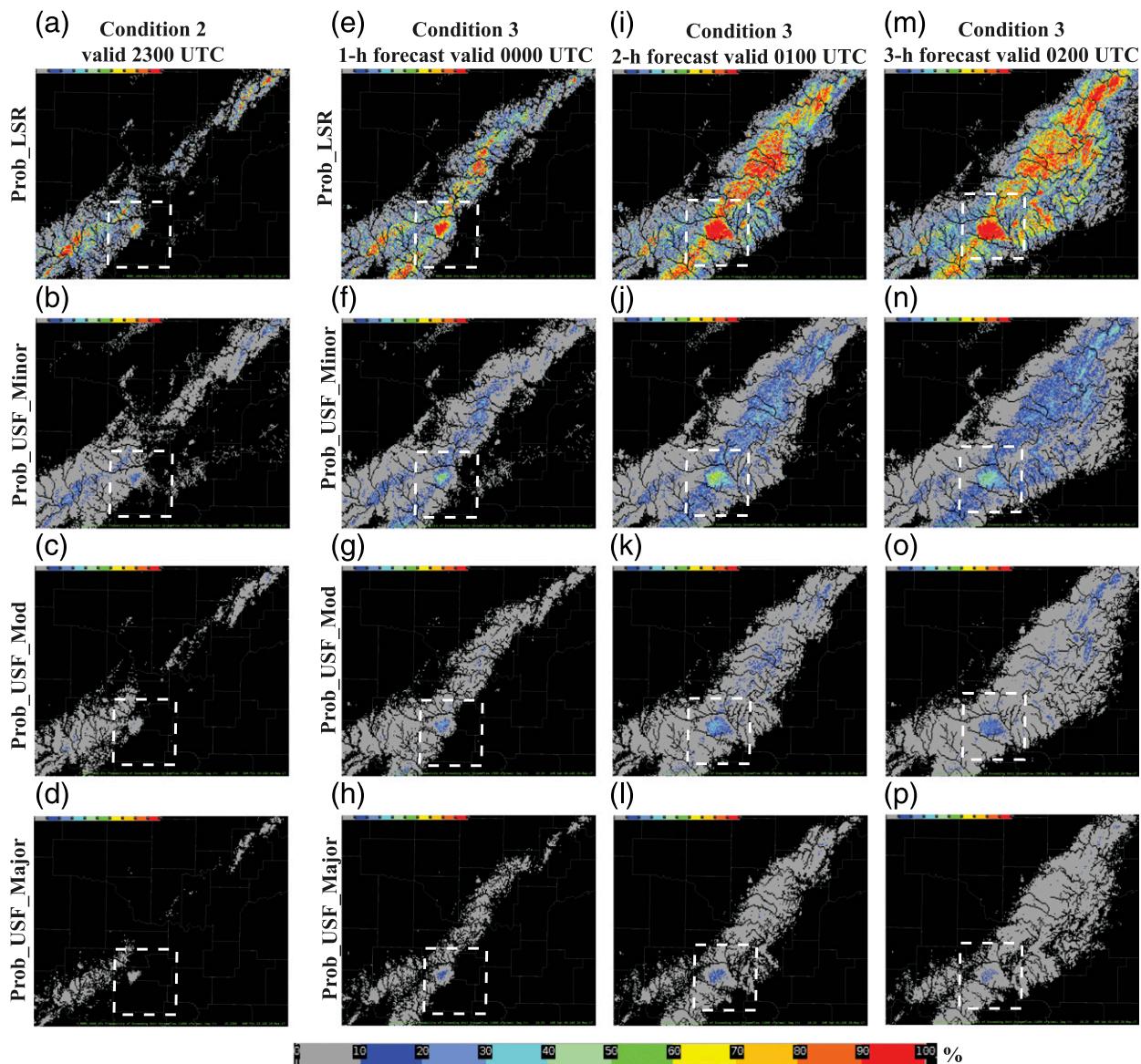
FIG. 8. Case 1: (a),(e),(i),(m) Prob_LSR (%), (b),(f),(j),(n) Prob_USF_Minor (%), (c),(g),(k),(o) Prob_USF_Mod (%), and (d),(h),(l),(p) Prob_USF_Major (%) from QPE-FLASH in (a)–(d) (condition 2; valid 2300 UTC 19 May 2017) and WoFS-FLASH in (e)–(p) (condition 3; valid 0000, 0100, and 0200 UTC 20 May 2017), respectively. The dashed white box in each panel is located over Murray and Carter Counties.

time only two participants expected to update information in current FFWs (Fig. 7).

2) CASE 2: 31 MAY–1 JUNE 2013

This case was completed by six of the nine participants, since it was only offered during the final 2 weeks of the experiment. The greatest difference in participants' attention and expected actions for Oklahoma County occurred in the first hour of the case (2300 UTC; Fig. 9). When viewing conditions 1 and 2 data during this hour, only one participant acknowledged the possible

threat in Oklahoma County (Fig. 9), and subsequently decided to communicate the flash flooding threat on NWSChat and in a social media post (Figs. 10 and 11). The other five participants focused their attention on rainfall occurring west and north of Oklahoma County (Fig. 9). However, when viewing condition 3, all participants acknowledged the potential flash flooding threat in Oklahoma County (Fig. 10). After seeing probabilities of up to 100% Prob_LSR, 60% Prob_USF_Minor, and 30% Prob_USF_Mod, five of the six participants reported expectations that they
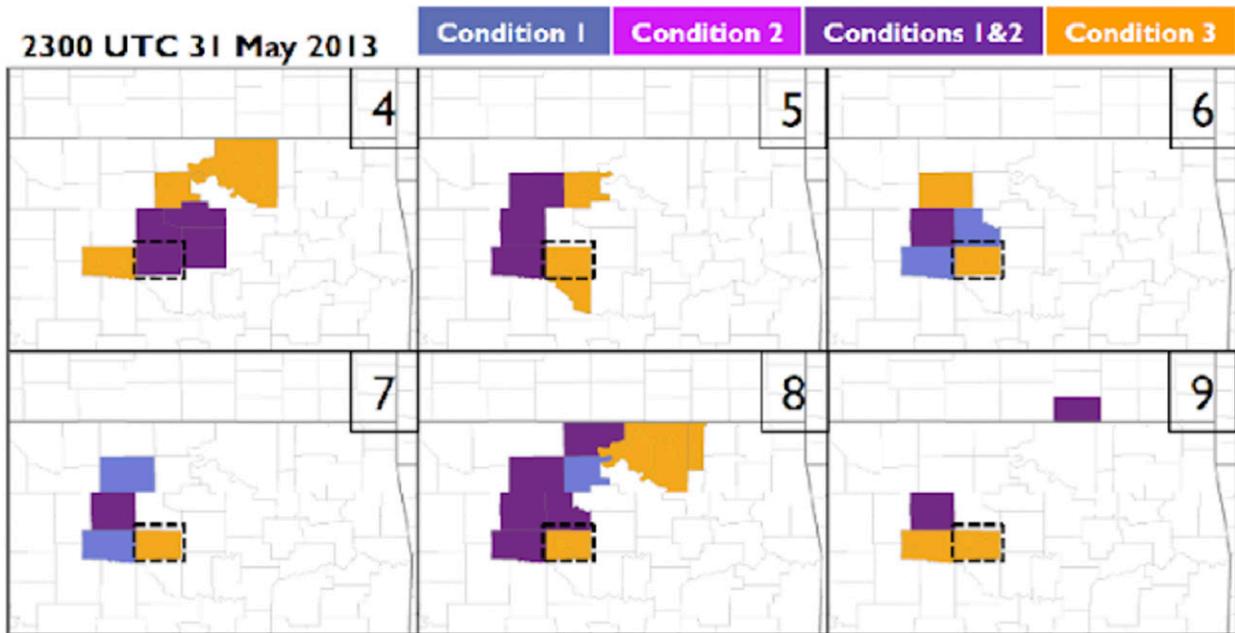
FIG. 9. Case 2: As in Fig. 6, but for six forecasters from 2300 UTC 31 May to 0100 UTC 1 Jun 2013 for the area highlighted in Fig. 8.

would issue a FFW or communicate the threat (Fig. 10). Two participants noted that the population center also influenced these expected actions. Although the 3-h probabilistic forecasts supported the likelihood of a flash flood event occurring, four participants noticed diminishing probability values in the third forecast hour (Fig. 11), which reduced confidence in some warning decisions. One participant said they ''wouldn't feel comfortable issuing a FFW at this point before it starts raining in the metro,'' and thus chose to only monitor the area at this time.

Five of six participants acknowledged the growing threat in Oklahoma County when looking at condition 1 in the second hour (0000 UTC; Fig. 10). Three of these participants reported seeing unit streamflow values increasing up to $3\,m^3\,s^{-1}\,km^{-2}$ ($274.5\,ft^3\,s^{-1}\,mi^{-2}$) and thought that FFWs were warranted. In contrast, one participant did not think the unit streamflow values were impressive, and therefore expected they would issue a flash flood advisory for Oklahoma County (Fig. 10). Two participants emphasized communications during this hour (Fig. 10), including one use of more impactful wording due to the concurrent tornado threat that could possibly dominate the scenario. As participants moved on to viewing condition 2 for 0000 UTC, they saw Prob_LSR values exceeding 90% and increasing probabilities for minor and moderate flash flooding. This guidance resulted in all participants paying attention to Oklahoma County, and two participants elevated their expected actions from condition 1 (Fig. 10).

At this same hour for condition 3, all participants continued to pay attention to Oklahoma County, and those that expected warning issuance in condition 2 also did so in condition 3. One participant continued to only
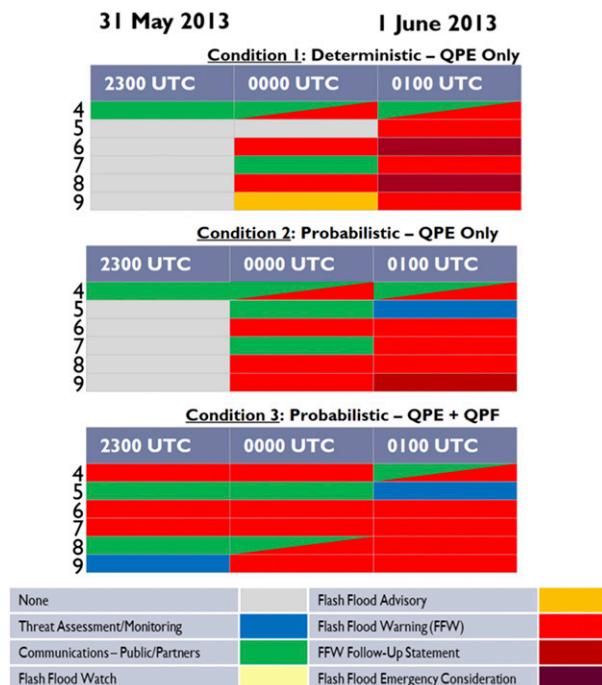

FIG. 10. Case 2: As in Fig. 7, but from six forecasters from 2300 UTC 31 May to 0100 UTC 1 Jun 2013 for the area highlighted in Fig. 9.
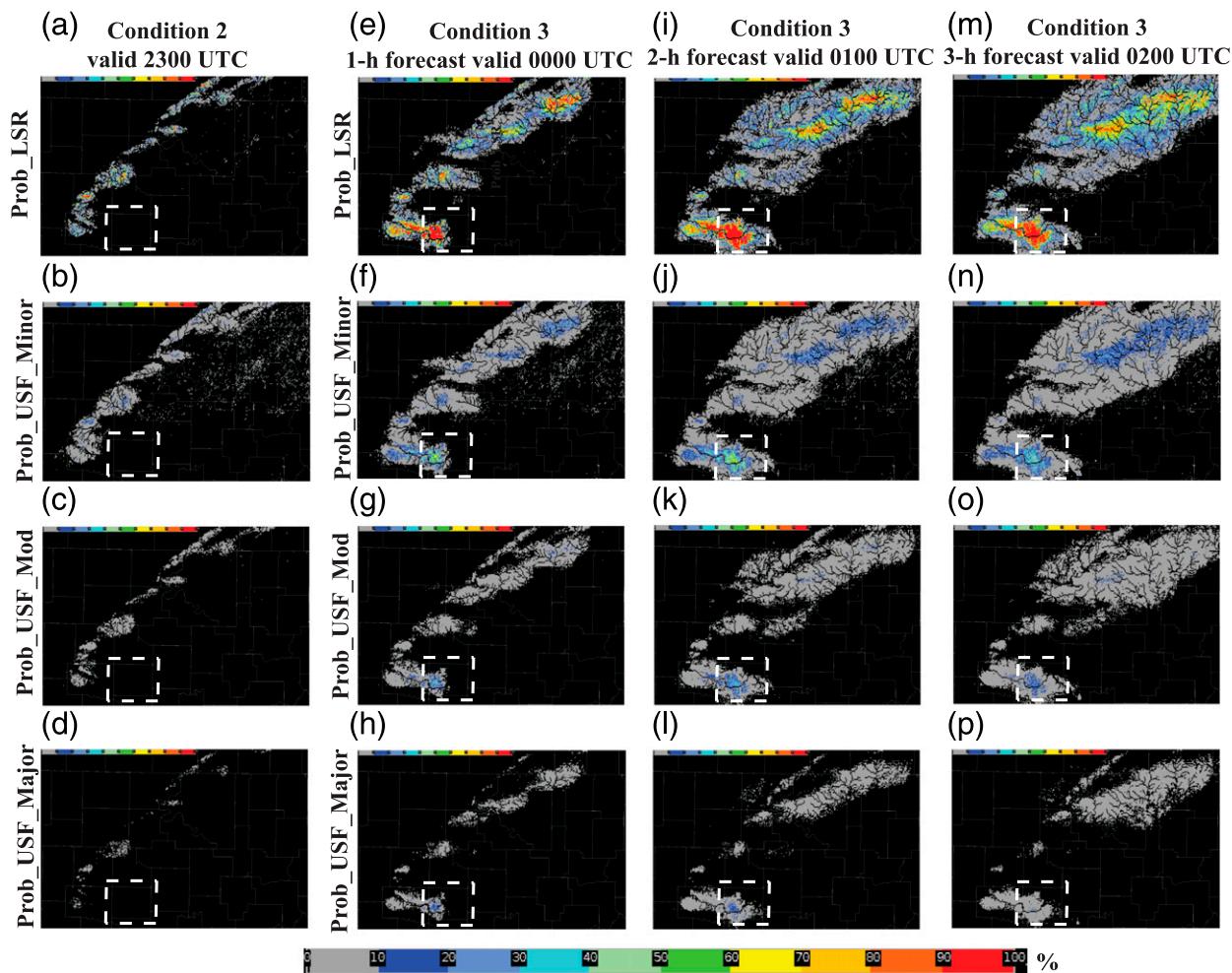
FIG. 11. Case 2: As in Fig. 8, but from QPE-FLASH in (a)–(d) (condition 2; valid 2300 UTC 31 May 2013) and WoFS-FLASH forcing in (m)–(p) (condition 3; valid 0200 UTC 1 Jun 2013), respectively. The dashed white box in each panel is located over Oklahoma County.

communicate the threat in condition 3, since the "threat still needs to be backed up by nearer term trends in convective evolution in observations" despite the 3-h forecast.

In the final hour of the case (0100 UTC), participants viewing condition 1 observed higher unit streamflow values exceeding $6 \, m^3 \, s^{-1} \, km^{-2}$ ($549 \, ft^3 \, s^{-1} \, mi^{-2}$), which led three participants who had not yet issued a FFW to now issue one (Fig. 10). Additionally, two participants that had already issued FFWs in condition 1 decided to issue updates with enhanced wording about the dangerous situation in Oklahoma City (Fig. 10). Most participants' perceptions of the flash flood threat shifted at 0100 UTC in condition 2 after viewing probability values lower than what the deterministic guidance led them to expect. This observation resulted in one participant believing that their expected action of warning issuance in condition 1 was now not warranted. By comparison, a different participant stated that their decision to issue a

FFW was reinforced in condition 2 because of relatively higher probability values in Oklahoma County compared to surrounding areas. Additionally, decisions to provide impactful wording in updates at 0100 UTC were no longer made. Only one participant expected to update a FFW to provide routine information (Fig. 10). After viewing condition 3 at 0100 UTC, participants saw decreasing probabilities over Oklahoma City in the 3-h forecast; thus, their attention now shifted to the county immediately to the south as the storm began to move into this location. Unlike in the previous conditions, no participants decided to issue an update (Fig. 10). One participant planned to let their FFW expire, while another contacted the emergency manager to let them know when to expect improvements in the flash flooding situation. The one participant who did not make a warning decision at all in condition 2 or 3 explained that the forecast probabilities "gave confidence that little
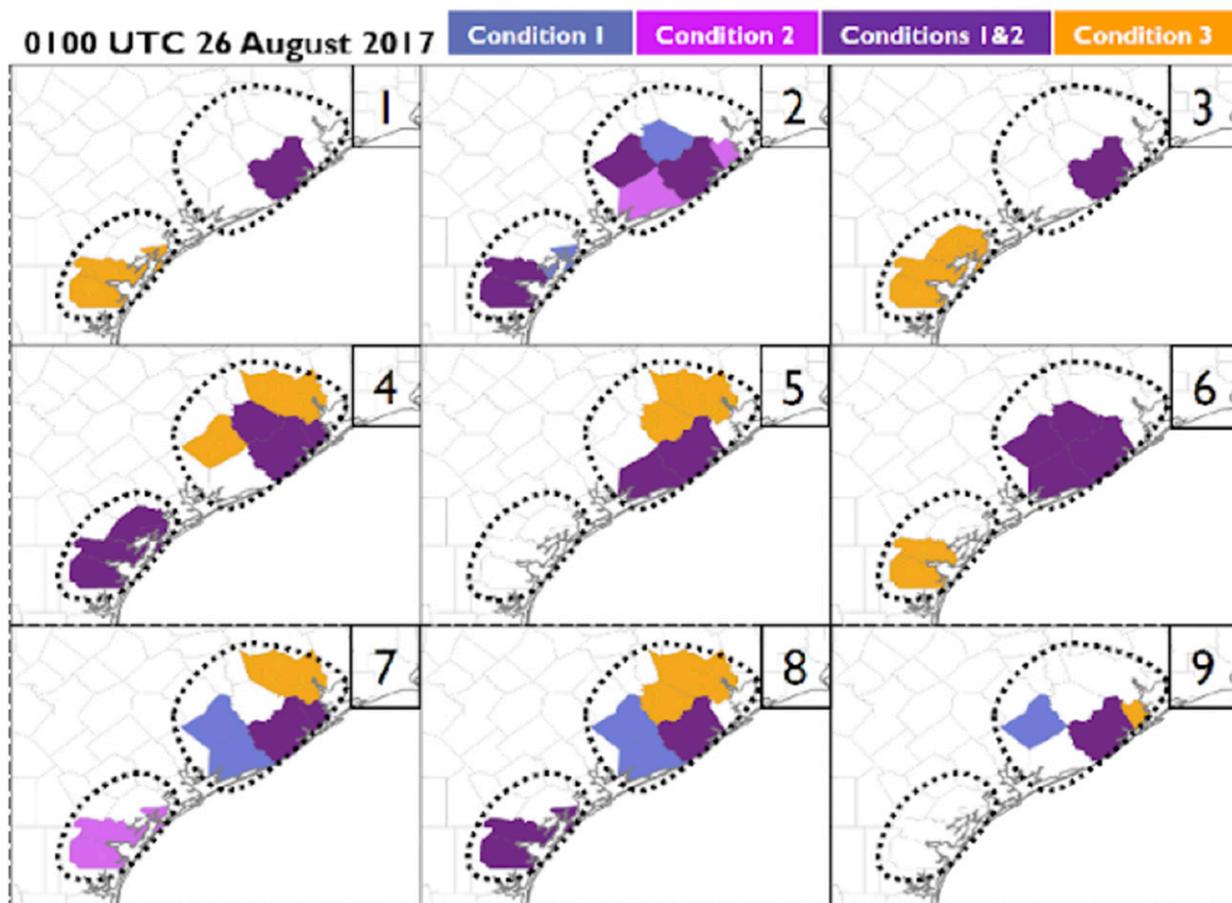
FIG. 12. Case 3: As in Fig. 6, but from 0100 to 0300 UTC 26 Aug 2017 for two areas (northern and southern) highlighted in Fig. 11.

additional impactful heavy rainfall will occur,'' and therefore did not provide convincing evidence that flash flooding would occur.

3) CASE 3: 26 AUGUST 2017

The decision-making processes in the Hurricane Harvey event occurred primarily within the first 3 h (0100–0300 UTC) of the case; thus, participants' attention and expected actions are discussed for this period only. The amount of attention given to the northern and southern areas and the types of actions reported varied most dramatically between all conditions in the first hour of the case (0100 UTC; Figs. 12 and 13). All participants focused their attention on the northern area for conditions 1 and 2 (Fig. 12) and reported numerous expected actions (Fig. 13). In condition 1, unit streamflow values up to $3 \, m^3 \, s^{-1} \, km^{-2}$ ($274.5 \, ft^3 \, s^{-1} \, mi^{-2}$) and QPE to FFG ratios ($>200\%$) prompted five participants to decide FFWs were necessary, while two participants decided to issue flood advisories (Fig. 13). Four of these participants also communicated the threat via NWSChat and social media. The remaining two participants only

monitored or communicated the threat in condition 1 (Fig. 13). When viewing condition 2 at 0100 UTC, participants similarly focused their attention on the northern area, but the probabilistic guidance caused a handful of participants to revise expected actions at this hour (Fig. 14). After seeing approximately 30% Prob_USF_Minor flash flooding, one participant downgraded their condition 1 expected actions of a FFW and communication to only monitoring the threat; however, another upgraded their action due to seeing high Prob_LSR values (Fig. 13). In both conditions 1 and 2 at 0100 UTC, participants paid little attention to the southern area. Only three participants in condition 1 and four participants in condition 2 either acknowledged the potential flash flooding threat associated with the eyewall making landfall, decided a FFW was necessary, or communicated the threat (Fig. 13).

When viewing condition 3 at 0100 UTC (Fig. 14), participants' attention and expected actions changed considerably compared to conditions 1 and 2. While seven of the nine participants continued to acknowledge the threat in the northern area, only one participant
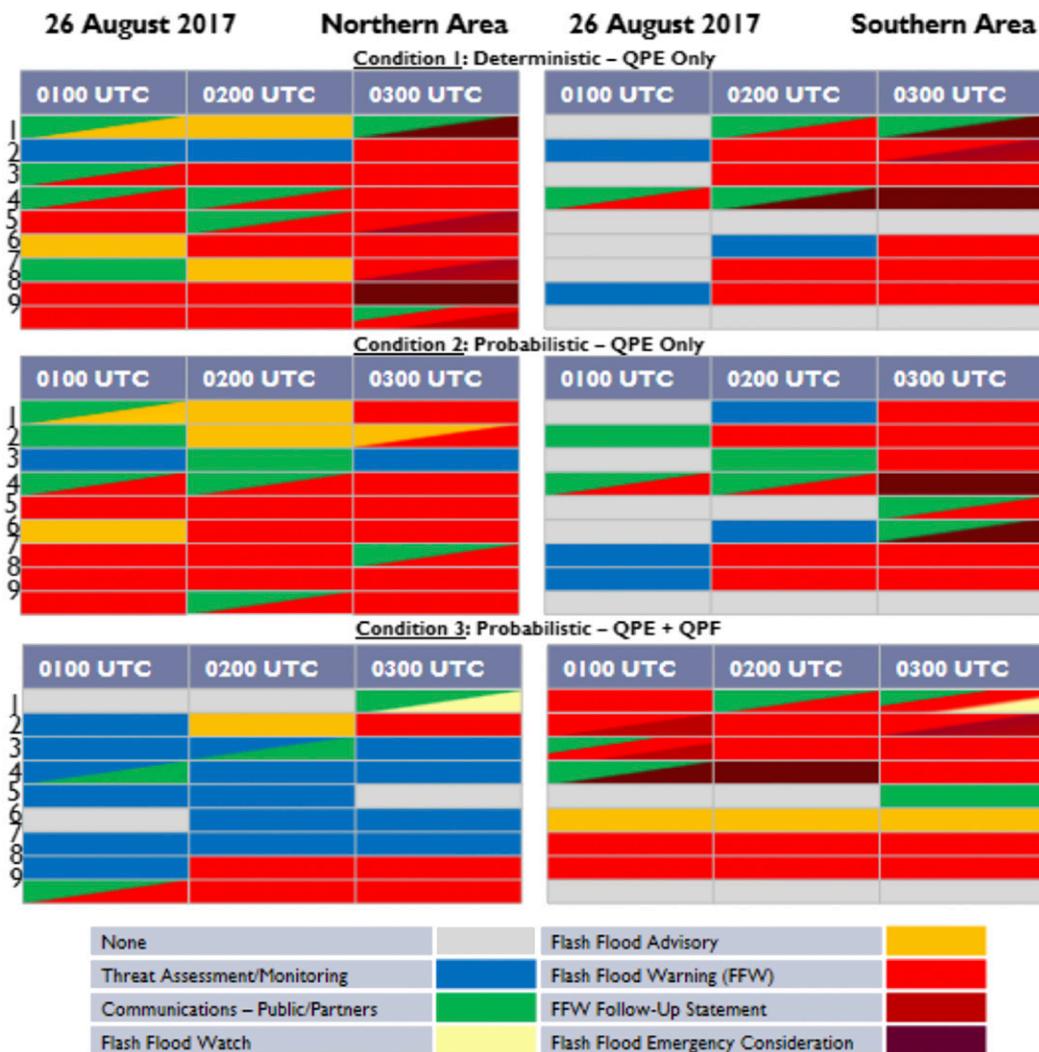
FIG. 13. Case 3: As in Fig. 7, but for two areas of interest (northern and southern areas) and from 0100 to 0300 UTC 26 Aug 2017 for two areas (northern and southern) highlighted in Fig. 12.

expected they would issue a FFW (Fig. 13). This change in expected actions was due to participants observing a decreasing trend in the areal extent of Prob_LSR values (Fig. 14). As concern in the northern area reduced while viewing the 0100 UTC condition 3 forecast, seven of the nine participants now attended to the southern area. Participants noted rainfall continuing throughout the 3-h forecast with elevated probabilities particularly in the first 2 h (e.g., 80% Prob_USF_Minor and 40% Prob_USF_Mod flash flooding along with 100% Prob_LSR values; Fig. 14). Based on these observations, five participants expected they would issue FFWs, one thought a flash flood emergency was required, and another issued an advisory. Three of these participants also decided to communicate the threat or issue an update (Fig. 13).

When viewing conditions 1, 2, and 3 data for 0200–0300 UTC, participants generally maintained their earlier perceptions of the flash flooding threat in the northern area. However, two participants upgraded their expected actions for this area in condition 1 at 0200 UTC (Fig. 13) after seeing QPE-to-FFG ratios up to 600% and unit streamflow values up to $6 \, \text{m}^3 \, \text{s}^{-1} \, \text{km}^{-2}$ ($549 \, \text{ft}^3 \, \text{s}^{-1} \, \text{mi}^{-2}$). Three participants also upgraded their expected actions at 0200 UTC in condition 2 (Fig. 13); however, these participants were not convinced of the flash flooding threat due to low Prob_USF_Minor values. At 0300 UTC in condition 1, participants viewed intense outer rainbands with continued high unit streamflow and QPE-to-FFG ratio values, which resulted in two participants citing the need for a flash flood emergency and three participants issuing updates to
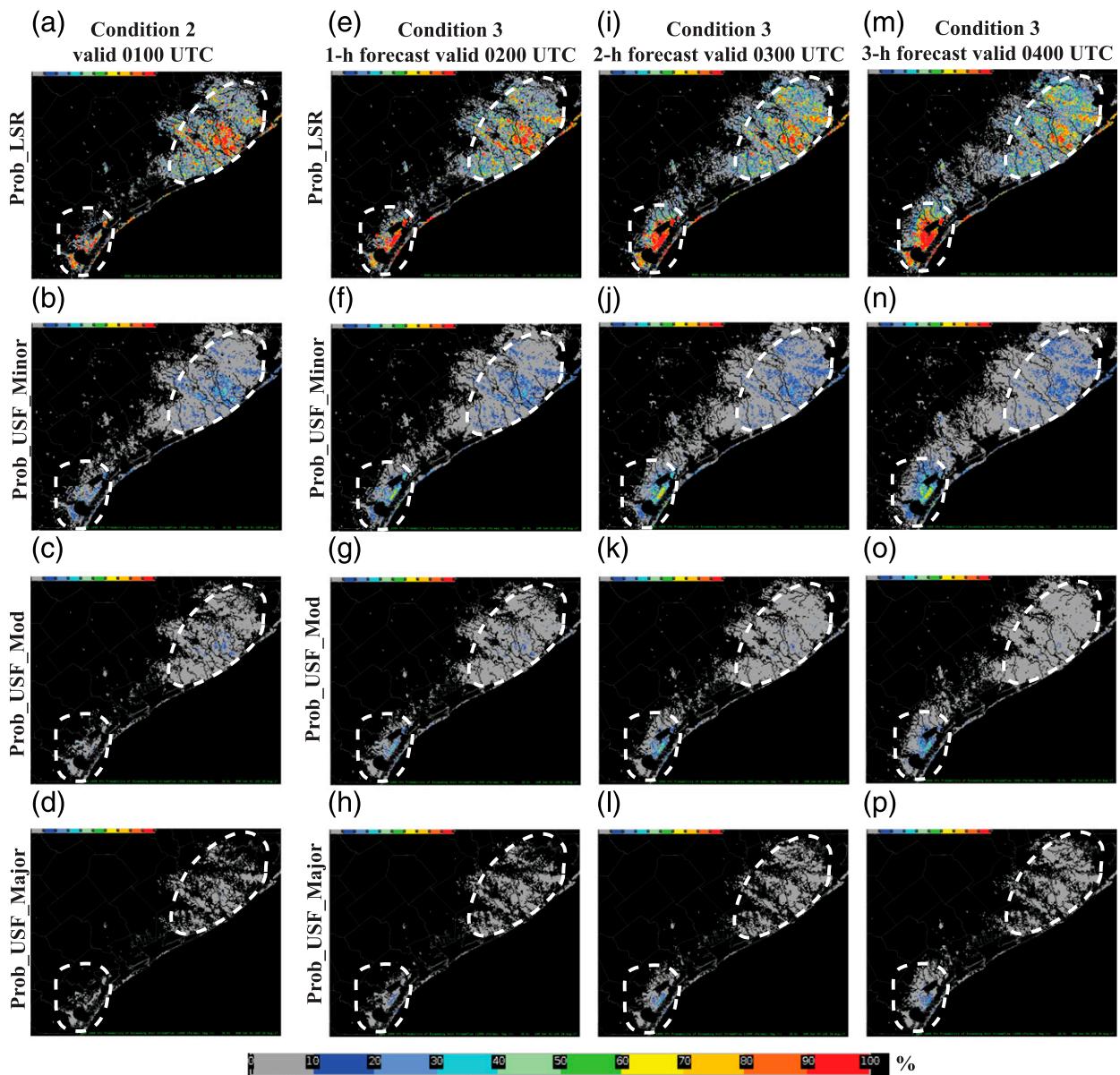
FIG. 14. Case 3: As in Fig. 8, but from QPE-FLASH in (a)–(d) (condition 2; valid 0100 UTC 26 Aug 2017) and WoFS-FLASH forcing in (m)–(p) (condition 3; valid 0400 UTC 26 Aug 2017), respectively. The dashed white lines in each panel outline the northern and southern areas.

FFWs (Fig. 13). Participants did not however feel that the low probabilities viewed in condition 2 at 0300 UTC were as suggestive of a flash flood threat. The probabilistic guidance therefore resulted in fewer participants elevating or updating warning information (Fig. 13). When assessing the northern area in condition 3 at 0200–0300 UTC, most participants did not believe FFWs were required due to the continued low Prob_USF_Minor, Prob_USF_Mod and Prob_USF_Major values; however, three participants made decisions to either issue a watch or a FFW due

to seeing that the flooding threat was pushing further inland (Fig. 13).

Participants' expected actions for the southern area at 0200 UTC were generally similar across the three conditions (Fig. 13). However, concern for the southern area at 0300 UTC was slightly more elevated for a few participants in conditions 1 and 2 compared to condition 3. This difference in concern was due to participants initially seeing high precipitation likely to cause localized major flooding as the eyewall moved inland. However, when viewing condition 3, participants were generally less

convinced of the major flash flooding threat due to the lower than expected probability values throughout the 0300 UTC 3-h forecast. Finally, two participants paying little to no attention to the southern area throughout the duration of the case in all conditions either did not notice this threat area until the final hour or focused their attention entirely on the northern area throughout the whole case.

c. Focus group summary

The focus group discussions first focused on participants' perceptions of the probability flash flood products used in conditions 2 and 3. All participants reported that the Prob_LSR values were "too hot" while the streamflow exceedance probabilities of Prob_USF_Major, Prob_USF_Mod, and Prob_USF_Minor did not appear sensitive enough and needed to be "tuned hotter." Although participants felt that improvements to these products need to be made, value was found when the Prob_LSR and Prob_USF threshold products were used in combination with one another. Using this approach, participants reported that the probability values influenced their decision making if they were previously undecided about whether a FFW was warranted or not when using the deterministic guidance. Some participants identified flash flooding threats using their own thresholds (e.g., 30%–40% values for moderate flash flooding), whereas others felt that they would develop mental thresholds and baselines after working through multiple cases. One challenge moving forward, as apparent from the analysis of participants' evaluations, will be establishing consistency in how probabilities are used to guide interpretation of hazardous threats. Participants demonstrated numerous instances during this study where the probabilistic guidance acted to either confirm expectations or reduce confidence in the occurrence of a flash flood event.

Participants were next asked how they envision the WoFS-FLASH guidance (condition 3) adding value to the warning decision process and communication of flash flooding. Responses indicated that participants were in consensus that the 3-h forecasts enabled them to direct their attention to threat areas that were not highlighted with the QPE-only driven data (conditions 1 and 2), which improved their overall situational awareness of the possible impending impacts. Some participants expect the WoFS-FLASH guidance to increase their confidence in the location of the impact rather than the actual magnitude of the impact. Additionally, participants anticipate the heightened situational awareness gained from WoFS-FLASH guidance would improve the decision support services they provide to both specific end users (e.g., emergency managers) and

to the general public. Some participants explained that this guidance would also support resource management, such as how many forecasters are needed on staff, whether the River Forecast Center needs to extend their operational hours for a certain event, and how workload should be best spread among forecasters.

Earlier detection of locations likely to experience flash flooding was evident in the evaluations obtained from the three cases. Many participants expected to take warning when using 3-h WoFS-FLASH guidance (condition 3), but this finding was not consistent among all forecasters. Therefore, when asked whether the probabilistic WoFS-FLASH guidance would positively impact FFW lead time, it was unsurprising to learn that participants were not in consensus. One participant explained that even if 3-h WoFS-FLASH guidance indicated the potential for flash flooding, they would not issue a FFW for a location if it was not yet raining. Other participants expect the impact on warning lead time to vary geographically and by storm type. Furthermore, participants' evaluations suggest that WoFS-FLASH guidance may shift attention away from ongoing threats, such that participants' expected actions indicated they were less likely to update current warnings and instead spend time assessing downstream threats.

The extent to which participants trusted or had confidence in the probabilistic products during the archived cases was also discussed. Participants reported a variety of reasons for increased confidence. Increased exposure to the products as participants worked through the three cases increased confidence. Additionally, participants were more confident using the probabilistic products during the convective cases rather than the tropical case, and when using the data to report expected actions of a watch versus warning issuance. Participants also reported increased confidence when assessing the probabilistic products closest to the time of model initialization, and when considering the trend in probabilities rather than the absolute values. Seeing geographical clustering of high probability values also improved confidence when interpreting a flash flooding threat. Although most participants identified instances when they felt confident using the probability products, one participant explained that learning to trust new tools is a "big deal" that would require viewing signals many times in operations before feeling confident enough to act on them.

5. Summary and future work

This study explores the application of the WoFS in a hydrologic context for probabilistic flash flood forecasting during the 2018 HMT-Hydro experiment using

three archived case studies. The participants analyzed the potential impacts of ingesting 0–3-h WoFS QPFs into the flash flood prediction process by utilizing several experimental probabilistic flash flood products developed within the FLASH hydrologic modeling system. The goal of this study was to identify how the probabilistic guidance products generated from the WoFS-FLASH system may enhance flash flood decision making and communication during operations.

The participants' evaluations of the coupled WoFS and hydrologic modeling system show promise of WoFS to enhance the decision-making process for flash flood events. Participants' threat assessment and monitoring phases started earlier with the addition of WoFS forcing (condition 3) compared to the QPE-only forcing (condition 2). In a number of instances, participants reported expected actions of warning issuance 1–3 h earlier when using the WoFS-FLASH guidance (condition 3) compared to the deterministic and QPE-only forcing guidance (conditions 1 and 2). The participants' expected actions also suggest that the WoFS-FLASH probabilistic products will result in earlier communication of flash flood threats to the public and partners. These findings suggest that WoFS-FLASH guidance will improve decision making during real-time flash flood events, and therefore motivate a subsequent study that will simulate real-time flash flood warning operations to examine measurable impacts on forecasters' decisions.

Some participants did not issue FFWs earlier, due to either a lack of confidence in the WoFS-FLASH system, not believing the probability values were indicative of a flash flooding threat, or being uncomfortable issuing warnings before rain fell in the forecasted threat area. Feedback from the participants also reveals biases in experimental probability products. The magnitude of the Prob_LSR products were generally perceived as higher than expected while the Prob_USF_Minor products were perceived as being too low for all three conditions and therefore warrants further tuning of the products. Therefore, numerous aspects of the model will need to be demonstrated prior to forecasters establishing trust with the output and using it to make real-life actionable decisions. Understanding how probabilities are calculated, and how they compare across products (e.g., what 50% means on LSR versus streamflow exceedance probabilities) will be important for ensuring that forecasters can effectively interpret and apply them during their decision-making processes.

The findings from the 2018 HMT-Hydro experiments are important for improving the probabilistic flash flood products, capturing initial impressions of the new probabilistic guidance, and forming expectations for how WoFS guidance may influence operational decision-making during flash flooding events within the FACETs paradigm. The lessons learned from the experiment provide a pathway to advance the science and application of WoFS PQPFs for operational hydrologic forecasts of flash flooding. The flash flood predictions are dependent not only on precipitation fallen on the ground but also on the characteristics of the underlying land surface. Accurately representing this coupled system is challenging and requires further interdisciplinary research collaborations between the meteorology and hydrology communities. Future work will explore useful probabilistic guidance products and optimization of those products for their eventual use in NWS flash flood watch and warning operations. Achieving progress in these areas will require not only basic research, but also collaborations between researchers, practitioners, emergency managers, and the public.

REFERENCES

Amengual, A., D. S. Carrio, G. Ravazzani, and V. Homar, 2017: A Comparison of ensemble strategies for flash flood forecasting: The 12 October 2007 case study in Valencia, Spain. *J. Hydrometeor.*, **18**, 1143–1166, https://doi.org/10.1175/JHM-D-16-0281.1.

Brown, J. D., D.-J. Seo, and J. Du, 2012: Verification of precipitation forecasts from NCEP's Short-Range Ensemble Forecast (SREF) system with reference to ensemble streamflow prediction using lumped hydrologic models. *J. Hydrometeor.*, **13**, 808–836, https://doi.org/10.1175/JHM-D-11-036.1.

Clark, R. A., J. J. Gourley, Z. L. Flamig, Y. Hong, and E. Clark, 2014: CONUS-wide evaluation of National Weather Service flash flood guidance products. *Wea. Forecasting*, **29**, 377–392, https://doi.org/10.1175/WAF-D-12-00124.1.

Davis, R. S., 2001: Flash flood forecast and detection methods. *Severe Convective Storms, Meteor. Monogr.*, No. 50, Amer. Meteor. Soc., 481–525.

Developmental Testbed Center, 2017a: Gridpoint Statistical Interpolation user's guide version 3.6. 158 pp., https://dtcenter.org/com-GSI/users/docs/.

——, 2017b: Ensemble Kalman Filter (EnKF) user's guide for version 1.2. 86 pp., http://www.dtcenter.org/EnKF/users/docs/index.php.

Devi, G. K., B. P. Ganasri, and G. S. Dwarakish, 2015: A review on hydrological models. *Aquat. Procedia*, **4**, 1001–1007, https://doi.org/10.1016/j.aqpro.2015.02.126.

Doswell, C. A., III, H. E. Brooks, and R. A. Maddox, 1996: Flash flood forecasting: An ingredients-based methodology. *Wea. Forecasting*, **11**, 560–581, https://doi.org/10.1175/1520-0434(1996)011<0560:FFFAIB>2.0.CO;2.

Dowell, D. C., and Coauthors, 2016: Development of a High-Resolution Rapid Refresh Ensemble (HRRRE) for severe weather forecasting. *28th Conf. on Severe Local Storms*, Portland, OR, Amer. Meteor. Soc., 8B.2, https://ams.confex.com/ams/28SLS/webprogram/Paper301555.html.

Duc, L., K. Saito, and H. Seko, 2013: Spatial–temporal fractions verification for high-resolution ensemble forecasts. *Tellus*, **65A**, 18171, https://doi.org/10.3402/tellusa.v65i0.18171.

Gourley, J. J., and Coauthors, 2017: The FLASH project: Improving the tools for flash flood monitoring and prediction across the United States. *Bull. Amer. Meteor. Soc.*, **98**, 361–372, https://doi.org/10.1175/BAMS-D-15-00247.1.

Hardy, J., J. Gourley, P. Kirstetter, Y. Hong, F. Kong, and Z. Flamig, 2016: A method for probabilistic flash flood forecasting. *J. Hydrol.*, **541**, 480–494, https://doi.org/10.1016/j.jhydrol.2016.04.007.

Jones, T. A., K. Knopfmeier, D. Wheatley, G. Creager, P. Minnis, and R. Palikonda, 2016: Storm-scale data assimilation and ensemble forecasting with the NSSL experimental Warn-on-Forecast system. Part II: Combined radar and satellite data experiments. *Wea. Forecasting*, **31**, 297–327, https://doi.org/10.1175/WAF-D-15-0107.1.

Lawson, J. R., J. S. Kain, N. Yussouf, D. C. Dowell, D. M. Wheatley, K. H. Knopfmeier, and T. A. Jones, 2018: Advancing from convection-allowing NWP to Warn-on-Forecast: Evidence in progress. *Wea. Forecasting*, **33**, 599–607, https://doi.org/10.1175/WAF-D-17-0145.1.

Martinaitis, S. M., and Coauthors, 2017: The HMT Multi-Radar Multi-Sensor Hydro Experiment. *Bull. Amer. Meteor. Soc.*, **98**, 347–359, https://doi.org/10.1175/BAMS-D-15-00283.1.

NWS, 2015: The May 31–June 1, 2013 tornado and flash flooding event. National Weather Service, https://www.weather.gov/oun/events-20130531.

Perica, S., and Coauthors, 2013: Version 2.0: Southeastern States (Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi). Vol. 9, Precipitation-Frequency Atlas of the United States, NOAA Atlas 14, 163 pp., http://www.nws.noaa.gov/oh/hdsc/PF_documents/Atlas14_Volume9.pdf.

Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97, https://doi.org/10.1175/2007MWR2123.1.

Rothfusz, L. P., R. Schneider, D. Novak, K. E. Klockow-McClain, A. Gerard, C. Karstens, G. Stumpf, and T. Smith, 2018: FACETs: A proposed next-generation paradigm for high-impact weather forecasting. *Bull. Amer. Meteor. Soc.*, **99**, 2025–2043, https://doi.org/10.1175/BAMS-D-16-0100.1.

Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., https://doi.org/10.5065/D68S4MVH.

Skinner, P. S., L. J. Wicker, D. M. Wheatley, and K. H. Knopfmeier, 2016: Application of two spatial verification methods to ensemble forecast of low-level rotation. *Wea. Forecasting*, **31**, 713–735, https://doi.org/10.1175/WAF-D-15-0129.1.

——, and Coauthors, 2018: Object-based verification of a prototype Warn-on-Forecast system. *Wea. Forecasting*, **33**, 1225–1250, https://doi.org/10.1175/WAF-D-18-0020.1.

Smith, T. M., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) severe weather and aviation products: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97**, 1617–1630, https://doi.org/10.1175/BAMS-D-14-00173.1.

Sorooshian, S., K.-L. Hsu, E. Coppola, B. Tomasseti, M. Verdecchia, and G. Visconti, 2008: *Hydrological Modeling and the Water Cycle: Coupling the Atmospheric and Hydrologic Models.* Springer, 291 pp.

Stensrud, D. J., and Coauthors, 2009: Convective-scale warn-on-forecast system: A vision for 2020. *Bull. Amer. Meteor. Soc.*, **90**, 1487–1499, https://doi.org/10.1175/2009BAMS2795.1.

——, and Coauthors, 2013: Progress and challenges with Warn-on-Forecast. *Atmos. Res.*, **123**, 2–16, https://doi.org/10.1016/j.atmosres.2012.04.004.

Sweeney, T. L., 1992: Modernized areal flash flood guidance. NOAA Tech. Memo NWS HYDRO 44, 35 pp., https://repository.library.noaa.gov/view/noaa/13498.

Vincendon, B., V. Ducrocq, O. Nuissier, and B. Vié, 2011: Perturbation of convection-permitting NWP forecasts for flashflood ensemble forecasting. *Nat. Hazards Earth Syst. Sci.*, **11**, 1529–1544, https://doi.org/10.5194/nhess-11-1529-2011.

Wang, J., and Coauthors, 2011: The Coupled Routing and Excess Storage (CREST) distributed hydrological model. *Hydrol. Sci. J.*, **56**, 84–98, https://doi.org/10.1080/02626667.2010.543087.

Wheatley, D. M., K. H. Knopfmeier, T. A. Jones, and G. J. Creager, 2015: Storm-scale data assimilation and ensemble forecasting with the NSSL experimental Warn-on-Forecast system. Part I: Radar data experiments. *Wea. Forecasting*, **30**, 1795–1817, https://doi.org/10.1175/WAF-D-15-0043.1.

Whitaker, J. S., T. M. Hamill, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation with the NCEP Global Forecast System. *Mon. Wea. Rev.*, **136**, 463–482, https://doi.org/10.1175/2007MWR2018.1.

Yussouf, N., and K. H. Knopfmeier, 2019: Application of Warn-on-Forecast system for flash-flood producing heavy convective rainfall events. *Quart. J. Roy. Meteor. Soc.*, **145**, 2385–2403, https://doi.org/10.1002/qj.3568.

——, D. C. Dowell, L. J. Wicker, K. H. Knopfmeier, and D. M. Wheatley, 2015: Storm-scale data assimilation and ensemble forecasts for the 27 April 2011 severe weather outbreak in Alabama. *Mon. Wea. Rev.*, **134**, 3415–3424, https://doi.org/10.1175/MWR3258.1.

——, J. S. Kain, and A. J. Clark, 2016: Short-term probabilistic forecasts of the 31 May 2013 Oklahoma tornado and flash flood event using a continuous-update-cycle storm-scale ensemble system. *Wea. Forecasting*, **31**, 957–983, https://doi.org/10.1175/WAF-D-15-0160.1.

Zhang, J., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97**, 621–637, https://doi.org/10.1175/BAMS-D-14-00174.1.